Face Segmentation under Unconstrained Scenes

Jing-Wein Wang

Institute of Photonics and Communications
National Kaohsiung University of Applied Sciences
E-mail: jwwang@cc.kuas.edu.tw

Abstract

In this paper, an efficient approach by combining the novel wavelet-based feature template, the support vector machine (SVM) classifier, and the wavelet entropy filtering is presented to robustly detect and segment human face image under complex background. Moreover, a face detection measure (FDM) criterion based on the distance between the expected and the detected eye-mouth triangle circumscribed circle areas is introduced to validate the performance of precise face segmentation.

1. Introduction

Facial recognition systems are computer-based security systems that are able to automatically detect and identify human faces. These systems usually start with the detection of face patterns. For the most part, the performance of face detection approaches reported in the literature has been measured on rectangle boundary, with each research site carrying out experiments using their own criterion and thus making meaningful comparisons and drawing conclusions impossible [1]. Amongst those trying to detect the face rectangle, a detected face is considered as valid if the overlaid frame window is not 20 percent bigger that the real face area and contain both eyes and mouth for the CFF method [2]. On the other hand, a face is considered correctly detected if the measured face position and size do not deviate from the true values for more than 30 percent for the RTFace method [3]. While no standard terms exist to define what a correct detection is, in the current work we devote our attention to how to segment a face precisely and therefore propose a face detection measure (*FDM*) criterion based on the distance between the expected and the detected eye-mouth triangle circumscribed circle areas to validate the performance of face detection.

The rest of this paper is organized as follows. In Section 2, the proposed face detection and segmentation approach will be introduced. Some experimental results will be demonstrated in Section 3 to corroborate the proposed approach. Section 4 concludes the paper.

2. Precise Face Segmentation

The algorithm's details of precise face segmentation are described as follows:

Step 1: Perform skin color detection [4] on input image (Fig. 1(a)) to locate region of interest (ROI), which is gotten by enlarging the skin region with a factor 1.2.

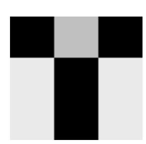
Step 2: Transform the ROI into overcomplete wavelet coefficients by using the D4 basis [5] as shown in (Fig. 1(b)).

Step 3: Set the initial feature template size to 24 × 24 pixels, exhaustively scan each ROI at different scales starting from the top left corner and sliding the template at two pixels increment horizontally and vertically until all ROIs have been covered. The ROI is scanned at 10 scales each a factor of 1.25 larger than the last.

Step 4: To locate the potential candidate (Fig. 1(c)), the rectangle feature of $W \times H$ pixels in the feature template can be computed as follows

$$[i] = \sum_{0 \le x < W, 0 \le y < H} | d_{LH}^{DWF}(x, y) |, \qquad (1)$$

where d_{LH}^{DWF} is the nonsubsampled version of discrete wavelet transform (*DWF*, discrete wavelet frames) [5] in the *LH* horizontal subband and $i \in \{1, ..., I\}$, $I = 9$ is the number of rectangles in the proposed feature template. At each location within ROI, check the decision rule of equation (2) that indicates whether the face template is satisfied or not.



1	2	3
4	5	6
7	8	9

([1] > [4]) && ([3] > [6]) && ([1] > [2] ‖ [3] > [2]) && ([5] > [4]) && ([5] > [6]) && ([8] > [7]) && ([8] > [9]) (2)

Step 5: Merge templates that can be counted as enclosing the same face since it is common to have multiple findings with small displacements horizontally and/or vertically.

Step 6: Normalize all the candidate regions – resizing to 24×24 pixels and using histogram equalization.

Step 7: Classify the face candidate using the *SVM* classifier [6]. If the class corresponds to a face, we draw a rectangle around the face in the output image (Fig. 1(d)).

Step 8: Normalize the detected face rectangle to size 40×60 pixels. An entropy-based smoothing filter is introduced to move the center from pixel to pixel in the rectangle of Fig. 1(d) to remove the coefficients located at outside the facial component regions. This continues until all pixel locations have been covered and facial component object are to be created for extraction. The entropy filtering of the pixels in the 3 × 3 (N = 3) neighborhood defined by the mask is given by the expression

$$Wavelet\_entropy = \frac{-1}{N^2} \sum_{x,y=0}^{N-1} \left| d_{LH}^{DWF}(x,y) \right| \log \left| d_{LH}^{DWF}(x,y) \right| . \quad (3)$$

According to anthropometry, we could perform the inter-orientation projection along the horizontal and vertical axes respectively to locate human eyes and mouth. Search for the centroid in the facial component region, we approach region segmentation by finding meaningful boundary based on point aggregation procedure. Choosing the center pixel of the component region is a natural starting point and grouping points to form the region of interest with paying attention to 4-connectivity would yield a clustering result, when no more pixels for inclusion in the region. After growing, the region centroid is relocated and therefore eyes, mouth, and face size are also known. An adaptation is finally carried out to mark the eyes and mouth positions and refine the bounding rectangle as an ellipse of fitting facial oval shape (Fig. 1(e)).

3. Experimental results

3.1. Photo detection

Our training set contains 16,000 faces and 22,000 non-faces which were extracted semi-automatically by our algorithm from various sources including internet and pictures taken by ourselves. The precise face detector is tested on the face images from the photos which differs from the face detector's training database. In order to define a correctly detected face in the case of precise face segmentation, we introduce the *FDM* criterion that a detected face is considered as valid if the detected eye-mouth triangle circumscribed circle area is within ±10 percent of the real face area and contains both eye-pair and mouth. The detection rate is the ratio between the number of successful detections and the number of labeled faces in the test set. The false alarm rate is the ratio between the number of false positive detection and the number of detected windows.

Fig. 2 shows example results obtained by our face detector on the test sets in comparison with the noted works based on convolutional neural network architecture (CFF) [2] and cascade boosting learning (RTFace) [3]. Among the three methods, our method does significantly well in locating facial components and contour, but with a few false alarms. The CFF algorithm performs as well with a low false rate, but has results of false negative for the dark face and incorrect detection for the masked face. The RTFace reports a result with higher false alarm rate than the previous twos. In the experiments, we obtain high detection rates with low false alarm, which are 93.69% with precision 84.63%, 90.38% with precision 88.68%, and 99.0% with precision 98.02% for the BioID (1521 images) [7], the Internet (100 images), and the Caltech (450 images) [8] test sets, respectively.

3.2. Live detection and recognition

As a case study, an integrated system for multiview face detection and recognition (*MFDR*) in complex scenes is constructed. The *MFDR* system comprises a real time face detector followed by a *SVM* recognizer. The *SVM* recognizer was trained and test on different sets of the live image database with 1000 positive examples (registered persons) and 1000 negative examples (intruders), 100 of each person, without using the face detector's ones. All the faces were collected in frontal view or near frontal view. The one-level decomposed wavelet subbands for the example image are all input to the *SVM* recognizer for face identification. We performed frontal-view and near frontal-view recognition based on the results of detection. In this scenario, the system processes a sequence of images to recognize the person and recall his name if he/she is recognized as a database subject; otherwise, the person is rejected as unknown. We perform face detection and recognition from video input lively based on rectangle face and ellipse face, respectively. Overall, the latter with 94% recognition rate and no false acceptance achieved a better performance than the former. In terms of speed, the execution time of the proposed live detector and recognizer is directly related to size and complexity of the images. For example, our system is operating at an average processing time 1.0 sec per live image with single face on a 3.0 GHz Pentium PC. The result shows that the performance of face recognition on this small scale problem (10 persons) is acceptable, therefore it may have potential applications such as door access control in a home security system.

4. Conclusions

We have presented in this paper a framework for precise face segmentation based on wavelet detector, without making any assumptions concerning the areas of the face pattern to analyze. The robustness of our system to varying poses and facial expressions as well as lighting variations was evaluated using real sets of difficult images. The satisfactory result for recognition has been obtained with the proposed ellipse face segmentation.

References

[1] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[2] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection", *IEEE Trans. Pattern Anal. Mach. Intell*,. vol. 26, pp. 1408-1423, 2004.

[3] B. Froba, A. Ernst, and C. Kublbeck, "Real-time face detection", *4th IASTED International Conference on Signal and Image Processing*, Kauai, 2002, pp. 479-502.

[4] St¨orring, M., *Computer Vision and Human Skin Colou*r, PhD thesis, Faculty of Engineering and Science, Aalborg University, Niels Jernes Vej 14, DK-9220 Aalborg, 2004.

[5] Mallat S., *A Wavelet Tour of Signal Processing*, Academic Press, 2nd Ed., 2001.

[6] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition", *Knowledge Discovery and Data Mining*, vol. 2, no. 2, pp. 1-47, 1998.

[7] http://www.humanscan.de/.

[8] http://www.vision.caltech.edu/html-iles/archive.html.

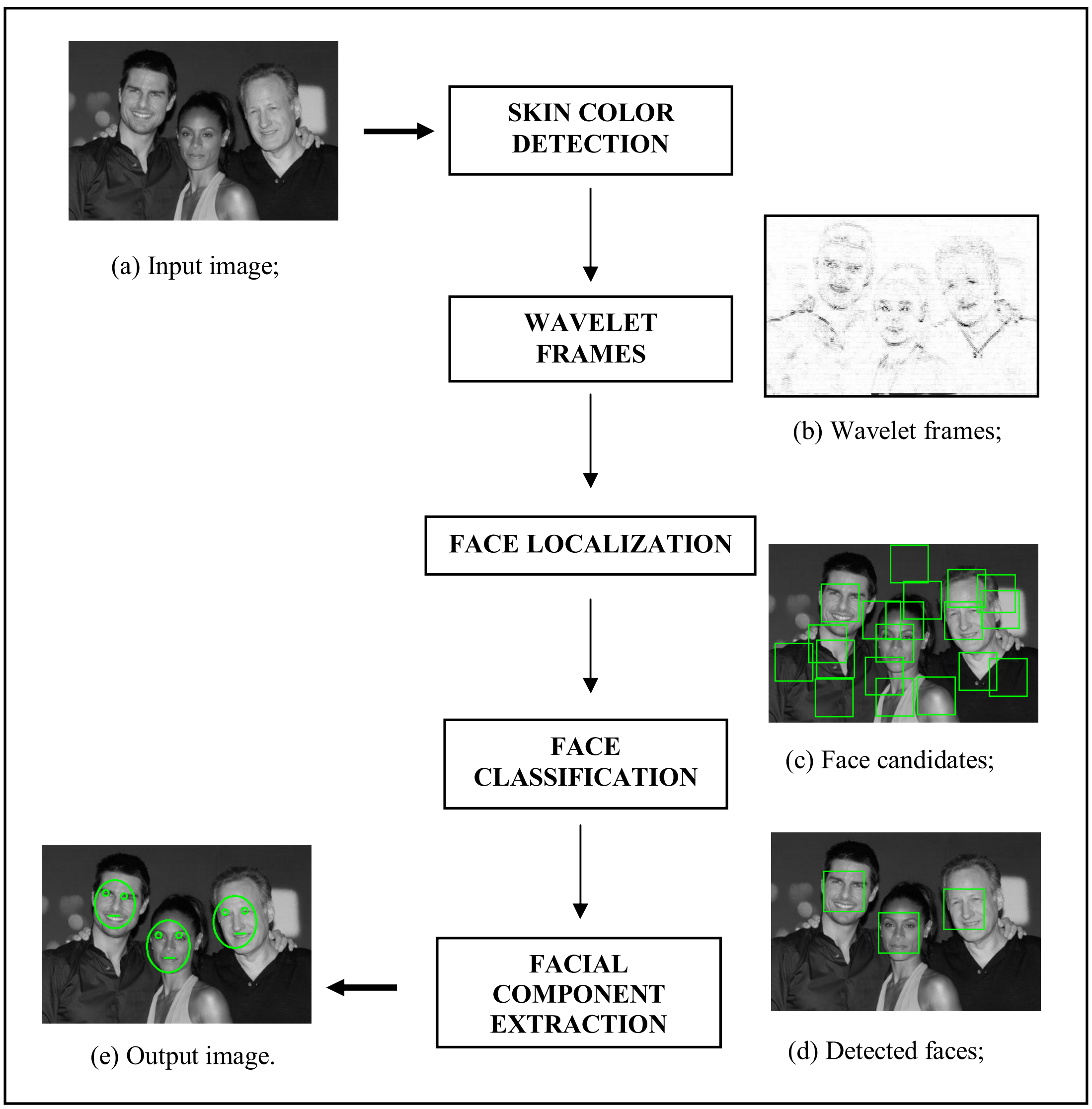


Fig. 1. Function block diagram of the proposed approach.

(a) (b) (c)

Fig. 2. Experimental results: (a) Our approach; (b) CFF; (c) RTFace.